A 7-ns 140-mW 1-Mb CMOS SRAM with Current Sense Amplifier

Katsuro Sasaki, Member, IEEE, Koichiro Ishibashi, Member, IEEE, Kiyotsugu Ueda, Kunihiro Komiyaji, Toshiaki Yamanaka, Member, IEEE, Naotaka Hashimoto, Hiroshi Toyoshima, Fumio Kojima, and Akihiro Shimizu

Abstract—A 7-ns 140-mW 1-Mb CMOS SRAM was developed to provide fast access and low power dissipation by using high-speed circuits for a 3-V power supply: a current-sense amplifier and a pre-output buffer. The current-sense amplifier shows three times the gain of a conventional voltage-sense amplifier and saves 60% of power dissipation while maintaining a very short sensing delay. The pre-output buffer reduces output delays by 0.5 ns to 0.75 ns. The 6.6-μm² high-density memory cell uses a new parallel transistor layout and phase-shifting photolithography. The critical charge that brings about soft error in a memory cell can be drastically increased by adjusting the resistances of poly-PMOS gate electrodes. This can be done without increasing process complexity or memory cell area. The 1-Mb SRAM was fabricated using 0.3-μm CMOS quadruple-poly and double-metal technology. The chip measures 3.96 mm × 7.4 mm (29 mm²).

I. INTRODUCTION

Today, workstations and computers are demanding faster memories in an effort to enhance system performance. To support this effort, many 1- and 4-Mb SRAM’s have achieved sub-10-ns access times [1]-[8]. Seven of these SRAM’s are BiCMOS SRAM’s [2]-[8] and only one is a CMOS SRAM [1]. However, these SRAM’s require large power dissipations of 275 to 1000 mW to attain their fast access times. Power reduction is essential in miniaturizing high-performance workstations and computers.

This paper reports a 1-Mb CMOS SRAM with 7-ns address access time and 140-mW power dissipation [9]. This SRAM has also reduced power dissipation to less than half of that in the previously reported 1-Mb SRAM’s. The fast access time and low power dissipation are achieved with new high-speed circuits for 3-V power supply: a current-sense amplifier and a pre-output buffer. The proposed current-sense amplifier has a large voltage gain for input voltages close to the supply voltage (V₉). This overcomes the gain reduction problem of conventional voltage-sense amplifiers operating with V₉-close input voltages. The V₉-close bit-line voltages are indispensable for stable low-voltage operation of SRAM’s. Although a current-sense amplifier with a different circuit configuration has been proposed and investigated with circuit simulation [10], this is the first paper reporting the successful implementation of a current-sense amplifier in a fabricated SRAM chip. The proposed pre-output buffer uses a simple configuration of an NMOS pull-up and a PMOS pull-down. Data output delay is reduced by pulsing these MOS transistors from the address transition detection (ATD) circuits. High-density and soft-error immunity are always important issues in a memory cell design. In this paper we will introduce a new parallel-transistor layout with very high density. We will also show that adjusting the resistances of polysilicon-PMOS gate electrodes can greatly improve the soft-error immunity.

II. HIGH-SPEED SENSE AMPLIFIER CIRCUIT

A fast sense-amplifier circuit is an important factor to a fast access time. In the 5-V design era, voltage-sense amplifiers were used to sense the small bit-line signals which were read out of a memory cell. As the size of these devices is reduced to the deep-submicrometer level, the supply voltage for these devices has to be reduced to less than 5 V to preserve the reliability. The growing density of the devices in a chip and the increasing operating frequency also make it necessary to reduce the supply voltage in order to decrease the power dissipation.

In 5-V supply SRAM’s, as shown in Fig. 1, NMOS transistors were generally used for the bit-line load transistors to obtain the appropriate input voltage levels for voltage-sense amplifiers. The bit-line voltage levels were around 3.5 V, which is lower than the supply voltage (5 V) by the threshold voltage of the load NMOS transistor. Conventional voltage-sense amplifiers [11] provide a large gain for such medium-input voltage levels. However, when the supply voltage is decreased to around 3 V, NMOS bit-line loads decrease the bit-line voltage to less than 2 V. As the high storage node voltage in a memory cell becomes equal to the bit-line voltage when the memory cell data are read, the noise margin of a memory cell is drastically degraded. This causes the degradation of cell stability and soft error immunity. Therefore, at a low power supply, the bit-line loads should be PMOS transistors so that bit-line voltages are again close to the supply...
voltage. Under these conditions, however, voltage-sense
amplifiers have poor voltage gain characteristics.

We have developed new current-sense amplifiers to
overcome this gain reduction at a low power supply.
Schematic diagrams of these amplifiers are shown in Fig.
2(a) and (b). A schematic diagram of a typical voltage-
sense amplifier is shown in Fig. 3 for comparison. In the
first of our current-sense amplifiers, shown in Fig. 2(a),
the input lines or common data lines are connected to the
source electrodes of PMOS transistors and the gate elec-
trodes of the PMOS’s are biased with the bias voltage
generator. Therefore, this is a PMOS bias type current-
sense amplifier. The bias voltage generator provides an
appropriate voltage to the PMOS’s so that they operate
near the saturation region. The NMOS’s have current-
mirror connections. In the second current-sense amplifier,
shown in Fig. 2(b), the gate electrodes of the NMOS’s are
biased with the bias voltage generator, which provides an
appropriate voltage to the NMOS’s so that they operate
near the saturation region. The PMOS’s have current-mir-
ror connections. This is an NMOS bias type current-sense
amplifier.

Fig. 4 shows the sensing circuitry with one of the pro-
posed current-sense amplifier (the PMOS bias type).
PMOS transistors are used for the bit-line loads. The op-
eration of the current-sense amplifier is explained as fol-
lows. When the word-line voltage is the ground level, the
same amount of current \( I_o \) flows from both bit lines (\( b_1 \)
and \( b_2 \)) to the amplifier. Therefore, the current that flows
through \( MP_1 \) to \( MP_3 \) in the amplifier is the same amount
(\( I_o /2 \)). When the word-line voltage rises to \( V_{CC} \), a small
amount of current \( \Delta I \) flows from one of the bit lines to a
cell. If, for example, cell current \( \Delta I \) flows from bit line
\( b_1 \) to the cell, the current flowing from bit line \( b_1 \) to the
amplifier is reduced to \( I_o - \Delta I \), while the current flowing
from bit line \( b_2 \) to the amplifier remains at \( I_o \). Consequently,
the current that flows through \( MP_1 \) and \( MP_3 \) is reduced to
(\( I_o - \Delta I \)/2), while the current that flows through \( MP_2 \)
and \( MP_4 \) remains at \( I_o /2 \). The current which flows through
\( MN_1 \) is \( I_o /2 \) because the current flowing through \( MP_2 \)
and \( MN_2 \) is the same, and \( MN_1 \) is the current mirror of \( MN_2 \).
Therefore, the load capacitance \( C_1 \) is discharged by \( MN_1 \), which draws more cur-
rent than \( MP_1 \) provides. In almost the same manner, the
load capacitance \( C_2 \) is charged up by \( MP_4 \), which pro-
vides more current than \( MN_4 \) draws. In this way, the vol-
age of terminal \( S_1 \) drops while the voltage of terminal \( S_2 \)
rises, and a voltage swing is obtained between \( S_1 \) and \( S_2 \).

In the operation of the current-sense amplifier, the bit-
line or common data-line voltage swing does not play an
important role in obtaining the voltage swing in the sense
amplifier output. This means that the current-sense am-
plifier can be used with a very small bit-line voltage
swing, which shortens the bit-line signal delay without
pulsed bit-line equalization. In the sensing circuitry,
shown in Fig. 4, a normally-on equalizer is used in the
read cycle to make the bit-line voltage swing small enough to attain a fast bit-line signal transition. Omitting the pulsed bit-line equalization is a power savings factor also.

We have investigated the characteristics of both the proposed current-sense amplifiers (Fig. 2) and a typical voltage-sense amplifier (Fig. 3) with circuit simulation. The simulation was done in actual memory circuits under conditions where the current of both amplifiers was 1.1 mA at a supply voltage of 3 V, and PMOS's were used for the bit-line loads to provide near-Vc, bit-line voltages. The load capacitance of C1 and C2 is 0.1 pF plus the gate capacitance of 20-μm-wide MOSFET's, which are the differential MOSFET's in the second sense amplifier. Since similar characteristics are obtained for both the PMOS bias type and NMOS bias type current-sense amplifiers, characteristics of the PMOS bias type are discussed in comparison to a voltage-sense amplifier. Fig. 5 shows the simulated voltage gain of the current-sense amplifier and the voltage-sense amplifier. The current-sense amplifier has almost three times the gain of the voltage-sense amplifier over a wide range of supply voltage. Fig. 6 shows the sensing delays for both the current- and voltage-sense amplifiers. Sensing delay is defined as the time interval from the crossing point of the input voltages to the point when the output voltage swing becomes 100 mV. The current-sense amplifier shows a shorter delay time by 30% than the voltage-sense amplifier. Fig. 7 shows the dependence of the sensing delay on the sense amplifier current. In the 1-Mb SRAM design, the current-sense amplifier shows a short delay of 1.3 ns with 1.1-mA sense amplifier current. The same delay is attained using the voltage-sense amplifier with as much as 2.8-mA sense amplifier current. The current-sense amplifier saves 60% of the power dissipation which is required for the voltage-sense amplifier to attain the same short delay time.

Fig. 5. Simulated voltage gain of the proposed current-sense amplifier and the conventional voltage-sense amplifier (Vc, is close to Vc).

Fig. 6. Simulated sensing delay of the current-sense amplifier and voltage-sense amplifier.

Fig. 7. Simulated dependence of sensing delay on sense amplifier current for the current-sense amplifier and voltage-sense amplifier.

III. FAST DATA OUTPUT CIRCUIT

In high-speed TTL SRAM's, data output delay occupies a significant part of the access time, 10% to 20% in general. Therefore, decreasing the data output delay is an important issue in a high-speed SRAM design. Towards this end, we propose a new fast data output circuit. A schematic diagram of the circuit is shown in Fig. 8. In the figure, Q1 and Q2 are ordinary output transistors to output data one and zero, respectively. Q1 and Q2 are turned off for a few nanoseconds with the output control signal φMA just before outputting data. The proposed pre-output buffer consists of an NMOS (Q3) connected to the Vc, and a PMOS (Q4) connected to the ground. These transistors are controlled by pulse signals generated with an ATD pulse generator. A fast data output is obtained by turning on Q3 and Q4 while Q1 and Q2 are turned off. We have evaluated the characteristics of the pre-output buffer by circuit simulation with a load capacitance Ci of 30 pF, which is a standard load capacitance for fast SRAM's.

The pull-up operation is shown in Fig. 9. When φMA is pulled down to the ground, terminals D1 and D2 are pulled up to Vc, Dn is pulled down, and output transistors Q1 and Q2 are turned off. During this period, Q3 and Q4 in the pre-output buffer are turned on with pulse signals φPT and φPB. In the pull-up operation, output terminal Dout is charged up through Q3 to the medium voltage level. Then, Q1 is turned on and data one is output. The output delay in the pull-up operation with the pre-output buffer is 0.75 ns less than the operation without the pre-output buffer. The pull-down operation is shown in Fig. 10. While Q1 and Q2 are turned off, Q3 and Q4 are turned on with pulse signals φPT and φPB. Output terminal Dout is discharged through Q4 to the medium voltage level. Then, Q2 is turned on and data zero is output. The output delay in the pull-down operation is shortened by 0.5 ns with the pre-output buffer.

Using the pre-output buffer, the peak current and noise on system bus are suppressed in zero-to-one or one-to-zero operation in addition to obtaining fast access, while additional charging and discharging current are required for zero-to-zero or one-to-one operation. Though the average system bus current may be increased, the peak current and noise on the bus in the worst case can be decreased. The operation of the pre-output buffer is based on the source-follower connection of an NMOS and a PMOS. This circuit configuration provides good pre-out-
put voltage levels, and at the same time it wastes little of the current passing through the two MOS transistors (which are turned on simultaneously). Fig. 11 shows the simulated pre-output voltage that the pre-output buffer provides. The pre-output voltage fits very well with a conventional TTL output reference of 1.5 V at low power supply voltages of 3 to 3.3 V. Fig. 12 shows the simulated shunt current through Q3 and Q4 in the pre-output buffer during operation. Because of the source-follower connection of Q3 and Q4, the shunt current is less than 0.1 mA at a supply voltage of 3 V. This current is negligible when compared to the total SRAM current.

The operating waveforms of the output circuits are shown in Figs. 13 and 14. Fig. 13 shows the pull-up operation which corresponds to Fig. 9. Before the address change, $D_p$ and $D_n$ are both in the high voltage level and the data output $D_{out}$ is zero because $Q1$ is turned off and $Q2$ is turned on. After the address change, $D_n$ is pulled down by $\phi MA$ (not shown in Fig. 13) and $Q2$ is turned off. While $Q1$ and $Q2$ are turned off, $Q3$ and $Q4$ are turned on by $\phi PT$ and $\phi PB$ and the output terminal $D_{out}$ starts to rise. Finally, $D_p$ is pulled down and the data one is output. Fig. 14 shows the pull-down operation that corresponds to Fig. 10. Before the address change, $D_p$ and $D_n$ are both in the low voltage level and the data output is zero because $Q1$ is turned on and $Q2$ is turned off. After the address change, $D_p$ is pulled up by $\phi MA$ (not shown in Fig. 14) and $Q1$ is turned off. While $Q1$ and $Q2$ are turned off, $Q3$ and $Q4$ are turned on by $\phi PT$ and $\phi PB$ and the output terminal starts to drop. Finally, $D_n$ is pulled up and the data zero is output.
IV. HIGH-DENSITY MEMORY CELL

It is desirable to have both a small cell area and excellent soft-error immunity. To obtain a very small cell area, we have used a new parallel transistor layout. As shown in Fig. 15, the word line is divided into two parts in this layout and the gate electrodes of the driver transistors are laid out parallel to the divided word lines. This configuration can minimize the useless area in a memory cell to achieve a small cell area of 2.0 \( \mu m \times 3.3 \mu m \) (6.6 \( \mu m^2 \)) with phase-shifting photolithography [12]. SEM micrographs of the memory cell, corresponding to the layouts in Fig. 15, are shown in Fig. 16. All the shapes are realized on processed and finished wafers.

It is well known that decoupling resistances in the cross-coupled connections of a full CMOS flip-flop memory cell have been used for obtaining single event upset (SEU) hardness [13], but this approach is not well suited to a high-density memory cell. To obtain a soft-error-immune high-density memory cell, we propose to use the gate electrodes of poly PMOS's as the decoupling resistances in a memory cell. We simulated the critical charge that brought about soft error with a circuit model shown in Fig. 17. The on current of the polysilicon PMOS was assumed to be about 0.2 \( \mu A \) for drain and gate voltages of 3 V. As shown in Fig. 17, we assumed device mismatches between the paired devices in a memory cell as follows: the difference of on currents in poly PMOS's is twice, the threshold voltage difference in driver NMOS's is 1.2 times, and the difference of the junction capacitances caused by the misalignment is 1.5 times. The alpha-particle-induced noise current used in the simulation is shown in Fig. 18. We used an exponential-like curve with a time constant of 30 ps. We simulated the voltage transitions of the two memory cell nodes after alpha particle hit during 1 \( \mu s \) for a supply voltage of 2 V. Examples of the simulated cell-node voltages are shown in Fig. 19. The upset of node voltages was recognized as the occurrence of soft error. The simulated critical charges for \( C_p \) of 0, 0.4, and 0.8 \( F \) are shown in Fig. 20. The results indicate that a few decades of kilohms of the decoupling resistance give a sharp increase in the critical charge and the required resistance is smaller for a large coupling capacitance \( C_p \). This implies that the soft error immunity is possibly improved by increasing the resistance of the poly-PMOS gate electrode to a few decades of kilohms. The resistance of the poly-PMOS gate electrodes can be adjusted into the desired range using ion implantation. With this approach, soft error immunity is enhanced with little added process complexity and no cell area increase. The increase in the
write pulse width for a decoupling resistance of 100 kΩ is less than 0.5 ns.

V. PROCESS TECHNOLOGY AND RAM CHARACTERISTICS

The 1-Mb SRAM was fabricated using 0.3-μm CMOS quadruple-poly and double-metal technology. The process technology is shown in Table 1. The first poly layer is used as gate electrodes of bulk MOS transistors. The second poly layer, which is actually tungsten polycide, is used as the memory cell ground lines. The third and fourth poly layers form polysilicon PMOS’s that are used as memory cell flip-flop loads; the third poly layer forms gate electrodes and the fourth poly layer is drain, source, and channel regions of the poly PMOS loads. The first metal is tungsten to attain excellent reliability, and the second metal is aluminum to obtain low resistivity.

A micrograph of the fabricated 1-Mb SRAM is shown in Fig. 21. The cell array is divided into two large blocks. Each block is further divided into 32 sections with each section including 64 × 256 cells. The peripheral circuits are laid out in the middle of the chip, which shortens the signal lines and reduces the signal delay. The chip measures 3.96 mm × 7.4 mm (29 mm²).

Address access time was measured at a supply voltage of 3 V at room temperature. The operating waveforms are shown in Fig. 22. A fast address access time of 7 ns was obtained. The power reduction in 1-Mb CMOS SRAM’s is shown in Fig. 23. Power dissipation is reduced from 482 mW in a 0.5-μm 1-Mb SRAM to 137 mW in a 0.3-μm 1-Mb SRAM. This is attained by the supply voltage reduction from 5 to 3 V, the short signal lines coming from the small chip size and the central peripheral circuit layout, and the newly developed low-power high-speed circuits which feature a current-sense amplifier and no
TABLE II
CHARACTERISTICS OF 1-Mb SRAM

<table>
<thead>
<tr>
<th>Configuration</th>
<th>256K × 4 bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power supply</td>
<td>3 V</td>
</tr>
<tr>
<td>Address Access Time</td>
<td>7 ns</td>
</tr>
<tr>
<td>Active Power</td>
<td>140 mW (100 MHz)</td>
</tr>
<tr>
<td>I/O Interface</td>
<td>TTL or CMOS</td>
</tr>
<tr>
<td>Chip Size</td>
<td>3.96 mm × 7.4 mm (29 mm²)</td>
</tr>
<tr>
<td>Memory Cell Size</td>
<td>2.0 µm × 3.3 µm (6.6 µm²)</td>
</tr>
</tbody>
</table>

pulsed bit-line equalization. The RAM characteristics are summarized in Table II.

VI. CONCLUSION

A 7-ns 140-mW 1-Mb (256K × 4) CMOS SRAM was reported. A fast access time and low power dissipation are achieved with newly developed high-speed circuits for a 3-V power supply. The circuits are a current-sense amplifier and a pre-output buffer. The current-sense amplifier shows three times the gain of a conventional voltage-sense amplifier and saves 60% of power dissipation in attaining a very short sensing delay. The pre-output buffer, which is shown to be suitable for a 3-V power supply, shortens the output delay by from 0.5 to 0.75 ns. A 6.6-µm² high-density memory cell was achieved using a new parallel-transistor layout and phase-shifting photolithography. It is also shown that the critical charge which brings about the soft error can be drastically increased by adjusting the resistances of poly-PMOS gate electrodes. Thus, this approach has great potential to improve soft-error immunity without increasing process complexity or memory cell area. The 1-Mb SRAM was fabricated using 0.3-µm CMOS quadruple-poly and double-metal technology. The chip measures 3.96 mm × 7.4 mm (29 mm²).

ACKNOWLEDGMENT
The authors wish to thank K. Shimohigashi, E. Takeda, T. Nishida, K. Seki, O. Minato, Y. Kawashima, T. Matsubara, S. Meguro, and S. Ikeda for their constant support and encouragement.

REFERENCES

Katsuro Sasaki (M’89) received the B.S. degree in electrical engineering and the M.S. degree in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1976 and 1978, respectively.

In 1978 he joined Semiconductor Division, Hitachi Ltd., Tokyo, Japan, where he was involved in the development of a 16-kb low-power CMOS static RAM and 16-kb, 64-kb, and 1-Mb high-speed CMOS static RAM’s. From 1985 to 1986 he worked on polysilicon TFT’s at the Massachusetts Institute of Technology, Cambridge. Since 1987 he has been working on the research and development of circuits and devices for submicrometer CMOS static RAM’s in the Central Research Laboratory, Hitachi Ltd., Tokyo, Japan. He is currently Manager of a research group on high-density CMOS RAM’s and high-speed CMOS/BiCMOS circuits.

Mr. Sasaki is a member of the IEEE Electron Devices Society and the Institute of Electronics, Information, and Communication Engineers of Japan.

Koichiro Ishibashi (M’86) was born in Tokyo, Japan, in 1958. He received the B.S. degree from Jia University, Tokyo, Japan, in 1980. He received the M.S. and the Ph.D. degrees in engineering from Tokyo Institute of Technology, Tokyo, Japan, in 1982 and 1985, respectively.

In 1985 he joined the Central Research Laboratory, Hitachi Ltd., Tokyo, Japan, where he has been engaged in the research and development of high-density CMOS static RAM’s.

Dr. Ishibashi is a member of the IEEE Electron Devices Society, the Japan Society of Applied Physics, and the Institute of Electronics, Information, and Communication Engineers of Japan.

Kiyotsugu Ueda was born in Nagasaki, Japan, in 1962. He graduated from the Nagasaki Technical High School, Nagasaki, Japan, in 1981.

He joined the Central Research Laboratory, Hitachi Ltd., Tokyo, Japan, in 1981 and has been working on research and development of CMOS static RAM’s.

Kunihiro Komiyagi was born in Kagoshima, Japan, in 1968. He graduated from Kagoshima Technical High School, Kagoshima, Japan, in 1986.

He joined the Central Research Laboratory, Hitachi Ltd., Tokyo, Japan, in 1986 and has been engaged in research and development of circuits for high-density CMOS static RAM’s.
Toshiaki Yamanaka (M'87) received the B.S. and M.S. degrees in electrical engineering from the University of Electro-Communications, Tokyo, Japan, in 1980 and 1982, respectively. In 1982 he joined the Central Research Laboratory, Hitachi Ltd., Tokyo, Japan. He has been engaged in the development of device and process technologies for high-density CMOS static RAM's. His research topic is to develop a polysilicon-TFT memory cell for 4- and 16-Mb SRAM's. His current interests also include the development of high-speed CMOS technology for logic and memory LSI's. Mr. Yamanaka is a member of the IEEE Electron Devices Society.

Fumio Kojima received the B.S. degree from Aoyamagakuen University in 1983. In 1983 he joined Hitachi VLSI Engineering Corporation, Tokyo, Japan, where he has been engaged in the research and development of submicrometer CMOS static RAM's.

Naotaka Hashimoto graduated from the Technical High School of Nagasaki, Nagasaki, Japan. He joined Hitachi Central Research Laboratory, Hitachi Ltd., Tokyo, Japan, in 1980, and has been working on the research and development of the MOS LSI process, especially on MOS device isolation. In 1983 he was a Visiting Researcher at the University of Waseda, Tokyo, Japan. Mr. Hashimoto is a member of the Japan Society of Applied Physics.

Hiroshi Toyoshima received the B.S. degree from Gunma University in 1986. In 1986 he joined Hitachi VLSI Engineering Corporation, Tokyo, Japan, where he has been engaged in the research and development of submicrometer CMOS static RAM's.

Akihiro Shimizu was born in Tochigi, Japan, in 1958. He received the B.S. degree in material physics from the University of Hiroshima, Hiroshima, Japan, in 1980. He was with Hitachi Microcomputer Engineering Ltd. from 1981 to 1984, where he worked on the MOS device technology. Since 1985 he has been working for Hitachi VLSI Engineering Corporation, Tokyo, Japan. His current research interests are in submicrometer MOS device technology for SRAM's and reliability physics such as hot-carrier effects. Mr. Shimizu is a member of the Japan Society of Applied Physics and the Institute of Electronics, Information, and Communication Engineers of Japan.